Measurement of the $t\bar{t}$ production cross section in the $\not{E}_T + jets$ channel with $2.2\ fb^{-1}$ of data

The CDF Collaboration
URL http://www-cdf.fnal.gov
(Dated: October 21, 2009)

In this note we describe the measurement of the $t\bar{t}$ production cross section in the final state characterized by missing transverse energy and high jet multiplicity ($\not{E}_T + jets$ final state), using a Neural Network to isolate the decay channel. The choice of this signature grants a high acceptance to general leptonic W decays, with a sizeable presence of $\tau + jets$ top pair decays, that are very difficult to isolate by means of the standard τ identification procedure. Moreover $\not{E}_T + jets$ $t\bar{t}$ decays provide complementary results with respect to standard lepton+jets, di-lepton, and all-hadronic top pair searches: using a particular choice of prerequisites cuts we can extract a signal sample orthogonal to those used by any other cross section analysis produced so far by the collaboration. This allows to obtain a measurement that is expected to have a strong impact on the combination of the results produced by the CDF experiment. The analysis reported in this note is based on 2.2 fb^{-1} of data collected up to August 2007 by a multi-jet trigger. After a set of clean-up cuts, a Neural Network is used to discriminate $t\bar{t}$ events from the background. After the requirement of at least one jet identified as a decay product of a b-quark, the cross section is extracted by means of a counting experiment on the sample of data whose Neural Network output is greater than 0.8. The amount of b-jets in the final selected sample due to background processes is estimated using parameterized probabilities of b-jet identification, measured directly from data.

The resulting $t\bar{t}$ production cross section is 7.99 ± 1.05 pb, assuming a top quark mass of $172.5\ GeV/c^2$.

I. INTRODUCTION

This note describes a measurement of the $t\bar{t}$ production cross section in $\bar{p}p$ collisions at $\sqrt{s} = 1.96$ TeV with the CDF detector at the Fermilab Tevatron. The measurement is performed looking for the top quark decay signature characterized by large jet multiplicity, large missing transverse energy due to neutrinos in the final state and at least one jet identified as a product of the decay of a b-quark (b-tagged jet). This signature will be referred to as $\not{E}_T + jets$ decay channel in the following. The search for this signature makes the measurement sensitive to W leptonic decays regardless of the lepton type and has a large acceptance with respect to the $W \to \tau\nu$ decays, very difficult to isolate with the standard τ identification tools. The search focuses on the missing transverse energy from the neutrino rather than on lepton identification, and thus it gives complementary and independent results with respect to the "lepton+jets" channel, where one W decays leptonically and the other into a pair of quarks; moreover, with appropriate cuts on the $\not{E}_T$, the measurement is also independent from the "all-hadronic" one, where both Ws decay hadronically. The impact on the combined cross section measurement is thus expected to be greatly enhanced.

The analysis starts with a series of clean-up cuts in order to perform a first raw selection of events with high $\not{E}_T$ and jet multiplicity. A Neural Network (NN) is then used to discriminate the signal from the background, mainly composed by QCD and $W/Z+$ heavy flavour jets events. A cut on the NN output selects the final sample, and the cross section measurement is extracted by means of a counting experiment on the number of jets identified as coming from the decay of a b-quark. The background in the final sample is estimated using parameterized b-jet identification probabilites derived from data. The resulting $t\bar{t}$ production cross section is 7.99 ± 1.05 pb, assuming a top mass of 172.5 GeV/c^2.

II. DATA SAMPLE & EVENT SELECTION

The analysis is based on an integrated luminosity of 2.2 fb^{-1} collected with the CDF detector between March 2002 and August 2007. The detector is described in detail in [1]. Data are collected with a multi-jet trigger which requires at least one calorimetric tower with $E_T \geq 10$ GeV at Level 1, the first of the three levels of the CDF trigger system; at Level 2 at least four clusters with $E_T \geq 15$ GeV and total transverse energy ΣE_T exceeding 175 GeV are required; finally, the Level 3 requires at least four jets with $E_T \geq 10$ GeV. As far as the Monte Carlo is concerned, to optimize the kinematical selection and determine its efficiency and systematics, we use samples generated with $M_{top} = 172.5$ GeV/c^2 using PYTHIA [2] and HERWIG [3], with different initial and final state radiation (ISR/FSR), parton density functions (PDFs) and color reconnection tunings to estimate the relevant systematic uncertainties. Generated events are reconstructed using a realistic simulation of the CDF detector and trigger system.

A. Event prerequisites

The following prerequisites are applied both to data and Monte Carlo samples before any kinematical selection:

- we require the events to belong to the good run list defined by the CDF data quality group with the requirement of fully operational silicon detectors, calorimeters and muon systems.

- we require at least one good quality vertex, formed with at least three tracks, to be present in the event and with a z coordinate within ± 60 cm from the geometrical center of the detector. We also require the vertex used for jet reclustering and then for the secondary vertex search to be close to the primary vertex found in the event, i.e. $|z_{jet} - z_{primvtx}|$ to be less than 5 cm;

- when dealing with Monte Carlo events, we perform a full simulation of the trigger path;

- we reject events with a good, $high - P_T$ reconstructed electron or muon to avoid overlaps with other top lepton+jets analyses [5];

- we clean up our sample by requiring events to have at least 3 $tight$ jets, i.e. jets with $E_T \geq 15$ GeV and $|\eta| \leq 2.0$ [4].

- We reject events with low missing transverse energy $\not{E}_T$ by requiring $\not{E}_T^{sig} \geq 3$ $GeV^{1/2}$, thus avoiding overlaps with top all-hadronic analyses [6].

The impact of these preliminary selections on data and inclusive Monte Carlo $t\bar{t}$ is shown in Tab. I along with the number of $t\bar{t}$ events expected in 2.2 fb^{-1} of data, assuming the theoretical production cross section of 7.45 pb,

CDF Run II Preliminary, L=2.2 fb^{-1}

Number of events	Data	MC_{incl}	Expected $t\bar{t}$ in 2.2 fb^{-1}
Good Run	14904280	4580599	15682
Trigger	10676105	2446149	8,374
$|Z_{vert}| < 60\ cm$	9781324	-	-
$|Z_{jet} - Z_{primvtx}| < 5\ cm$	9216501	2442174	8361
N_{vert} good quality ≥ 1	9216500	2442174	8361
N tight leptons $= 0$	9204642	2199776	7531
NJets ≥ 3	8786019	2199302	7,529
$\not{E}_T^{sig} \geq 3\ GeV^{1/2}$	138527	396924	1359
Out of which:			
with NJets$= 3$	44310	9087	31
with NJets≥ 4	94217	387837	1328
with NJets≥ 4 and $\Delta\phi_{min} > 0.4$	20043	207381	710

TABLE I: Effect of prerequisites cuts.

corresponding to a top mass $M_{top} = 172.5\ GeV/c^2$ [7]. Tab. II shows the number of events in the data sample and the $t\bar{t}$ contamination expected from Monte Carlo for different tight jet multiplicities.

CDF Run II Preliminary, L=2.2 fb^{-1}

Number of Events	3 jets	4 jets	5 jets	6 jets	7 jets	8 jets
Data	44310	52691	22760	9871	2714	660
$t\bar{t}$ MC	9090	107938	152740	87342	30074	7789
Exp. $t\bar{t}$ in 2.2 fb^{-1}	31	371	524	300	103	27
Exp. Contamination (%)	0.07	0.70	2.31	3.04	3.81	4.05

TABLE II: Expected signal contamination for different jet multiplicities.

III.   BACKGROUND PREDICTION

The cross section measurement is performed counting the number of jets identified by the SECVTX [8] algorithm as coming from the decay of a b-quark. The algorithm looks for jets with secondary vertices displaced from the primary one and classifies the jets as *positive* b-tagged if the secondary vertex is on the same side as the jet cone axis, as *negative* b-tagged otherwise. The former are most likely coming from the decay of a b-quark, the latter arise mainly from tracks mismeasurements. In order to derive a cross section measurement from the final tagged sample we need to find an estimate of the number of positive b-tagged jets yielded by background processes.

The basic idea of our background prediction method rests on the assumption that b-tag rates for $t\bar{t}$ signal and background processes show differences that are due to the different properties of the b-jets produced by the top quark decays compared to the b-jets arising from QCD and vector boson plus heavy flavour production processes. In this hypothesis, parameterizing the b-tag rates as a function of some chosen jet characteristics, in events depleted of signal contamination, allows to predict the number of b-tagged jets from background processes present in a given selected sample. To ensure low $t\bar{t}$ contamination in the sample used to study the b-tag rates, we select events with exactly 3 jets (see Tab. II) and define the b-tagging probability as the ratio of the number of positive SECVTX tagged jets to the number of taggable jets in the sample, where we define as taggable a jet with at least two good SECVTX tracks. The per-jet b-tagging probability has been studied as a function of several jet and event variables, and is found to depend mainly on jet characteristics such as E_T, the number of good quality tracks contained in the jet cone N_{trk}, and the $\not{E}_T$ projection along the jet direction $\not{E}_T^{prj}$. Figure III shows both the positive and negative tagging rates dependence on the set of variables chosen to parameterize the tagging probability.

These dependencies parameterized on 3-jets events are used to build a 3-dimensional b-tag rate matrix capable of assigning to a jet originated by background processes, given its characteristics, a probability to be identified as a b-jet.

This allows to calculate the number of background b-tags expected in a given data sample by evaluating the probability of tagging its jets.
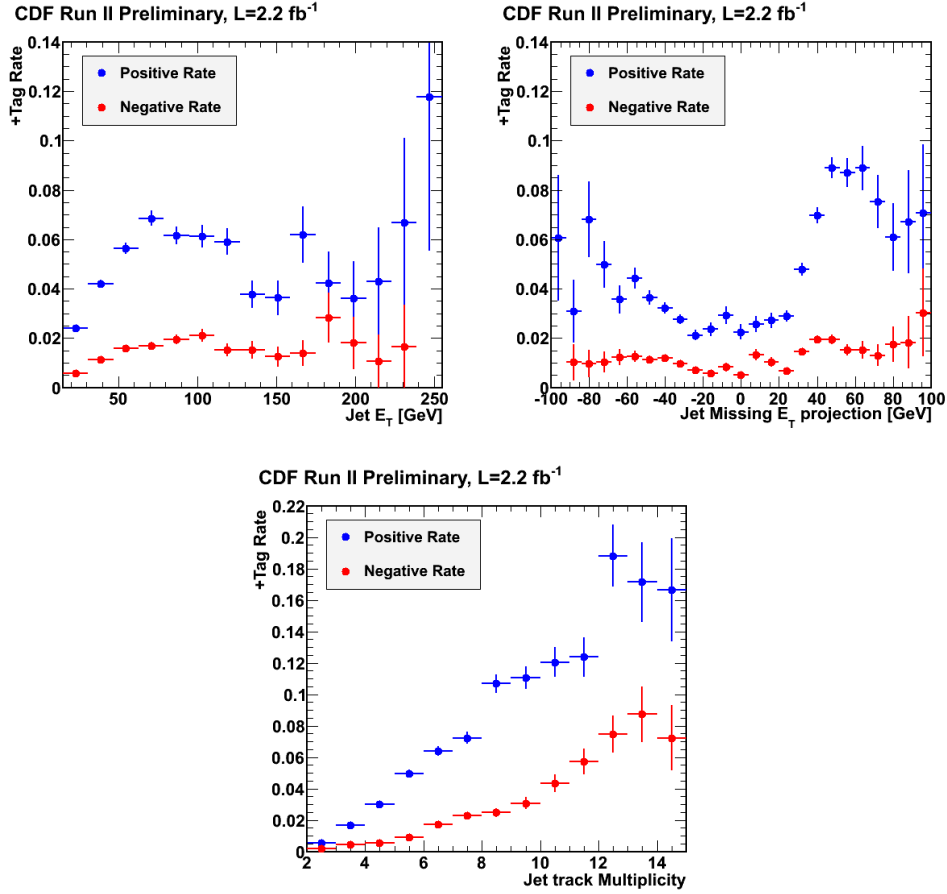
FIG. 1: b-jet tagging rate as a function of jet E_T, jet number of tracks and missing energy projection along the jet direction.

A. Iterative correction for the presence of top

Before using the matrix to evaluate the predicted number of positive b-tags due to background processes in a given data sample, we have to ensure that it contains a negligible $t\bar{t}$ component. Otherwise the expected number of b-tags provided by the matrix parameterization will not refer only to background events, as it will receive a contribution from $t\bar{t}$ events in the sample.

We can correct for this effect by removing the $t\bar{t}$ contribution in the pre-tag sample in order to have a background-only determination of the number of expected b-tags. To do so, we iteratively correct the number of expected b-tags in the sample as follows:

$$N'_{exp} = N^{fix}_{exp} \frac{N_{evt} - N^{t\bar{t}}_{evt}}{N_{evt}} = N^{fix}_{exp} \frac{N_{evt} - (N_{obs} - N_{exp})/\epsilon^{ave}_{tag}}{N_{evt}} \tag{1}$$

where N^{fix}_{exp} is the number - fixed during the iterative procedure - of expected tags coming from the tag rate parameterization before any correction, N_{evt} is the number of events in the pre-tagging data sample used to determine N^{fix}_{exp} through the tag matrix prediction, ϵ^{ave}_{tag} is the average tagging efficiency, defined as the Monte Carlo ratio between the number of positive b-tagged jets and the number of events in the pre-tag sample, and $N^{t\bar{t}}_{evt}$ is the $t\bar{t}$ contamination in the pre-tagging sample, estimated as $(N_{obs} - N_{exp})/\epsilon^{ave}_{tag}$. The iterative procedure stops when the difference $|N'_{exp} - N_{exp}| \leq 1\%$.

IV. KINEMATICAL SELECTION

In order to enhance the signal to background ratio in our final sample, we use a Neural Network (NN), trained to discriminate $t\bar{t} \to \not{E}_T + jets$ signal events from background. The NN is built using the class *TMultiLayerPerceptron* available in ROOT [9].

For what concerns training samples, as background we use all data collected with the multi-jet trigger and passing the prerequisites discussed in Sec. II A; additionally, we require the presence of at least 4 *tight* jets in the event to perform the training on a sample completely uncorrelated with the one we used to determine the background parameterization. Finally, in order to further clean up our sample we remove events with low angle between jets and $\not{E}_T$ (mainly due to energy mismeasurements and difficult to model) using an additional cut on the minimum difference in ϕ-coordinates between $\not{E}_T$ and jets in the event, requiring $\Delta\phi_{min}(\not{E}_T, jets) > 0.4$. For signal we use the same amount of events passing the same requirements of data, taken randomly from the available Monte Carlo sample. We use as inputs for the network the following kinematical variables, normalized with respect to their maximum values:

- E_{T1}, the transverse energy of the leading jet;

- $\Delta\phi_{min}(\not{E}_T, jet)$, the minimum difference between the $\not{E}_T$ and each jet in the event in the ϕ coordinates;

- $\not{E}_T^{sig}$, the $\not{E}_T$ significance of the event, defined as $\not{E}_T/\sqrt{\Sigma E_T}$;

- the energy-related variables $\sum E_T$, $\sum E_T^3$ and the Centrality of the energy flux, defined as $C = \frac{\Sigma_i E_{Ti}}{\sqrt{s}}$;

- the topology-related variables Sphericity $S = \frac{3}{2}(Q_1+Q_2)$ and Aplanarity $A = \frac{3}{2}Q_1$, where Q_i are the eigenvalues of the momentum tensor.

Fig. 2 shows the output of the NN and the corresponding background prediction from the tag matrix along with the expected contribution from $t\bar{t}$ signal for all data events with at least four jets and for events with exactly three, four and five tight jets. Matrix predicted b-tagged jets for bins with a considerable amount of signal contamination have been corrected according to the iterative procedure described in Sec. III A. Results are quite good over all the Neural Network spectrum, and in the high Neural Network output region we can see that the tagging matrix predictions are not sufficient to justify the number of observed tags, while the agreement is good if we add the amount of tags coming from the expected $t\bar{t}$ signal contribution. This is both a confirmation of the effectiveness of the method we used to estimate the background and a check of the correct behaviour of the Neural Network we trained.

V. SOURCES OF SYSTEMATIC ERRORS

The production cross section we want to measure is given by:

$$\sigma(p\bar{p} \to t\bar{t}) \times BR(t\bar{t} \to \not{E}_T + jets) = \frac{N_{obs} - N_{exp}}{\epsilon_{kin} \cdot \epsilon_{tag}^{ave} \cdot L} \qquad (2)$$

where N_{obs} and N_{exp} are the number of observed and corrected matrix-predicted tagged jets in the selected sample, respectively; ϵ_{kin} is the trigger, prerequisites and Neural Network selection efficiency measured on inclusive Monte Carlo $t\bar{t}$ events; ϵ_{tag}^{ave}, defined as the ratio of the number of positive tagged jets to the number of pre-tagging events in the inclusive $t\bar{t}$ Monte Carlo sample, gives the average number of b-tags per $t\bar{t}$ event. Finally, L is the luminosity of the dataset used.

As far as systematic uncertainties are concerned, ϵ_{kin} is affected by the uncertainty arising from the particular choice of the Monte Carlo generator used to estimate the kinematical efficiency, the set of parton density functions used, the modeling of color reconnection effects and of initial and final state radiation. All these uncertainties are estimated comparing the effects of the selection cuts on samples differing for Monte Carlo generator, PDFs, color reconnection models and amount of initial and final state radiation.

In order to account for the jet response systematic in the cross section measurement, we vary the corrected jet energies within $\pm 1\sigma$ of their corresponding systematic uncertainty and recalculate the ϵ_{kin} after these variations.

ϵ_{kin} is also affected by the simulation of the trigger requirements on Monte Carlo events and an uncertainty is derived comparing the trigger turn-on curves on data and Monte Carlo samples.

The systematic uncertainty on the average number of b-tags per $t\bar{t}$ event is obtained varying it from its central value within the $\pm 1\sigma$ range, while the uncertainty on the number N_{exp} of matrix-predicted tagged jets is is calculated by comparing the number of b-tags yelded by the tagging matrix application to the actual number of observed positive SECVTX tags in a control sample depleted of signal contamination.

We take into account also the systematic error on luminosity, coming from the uncertainty on the inelastic $p\bar{p}$ cross section and from detector effects, and the systematic error on the primary vertex selection cuts.
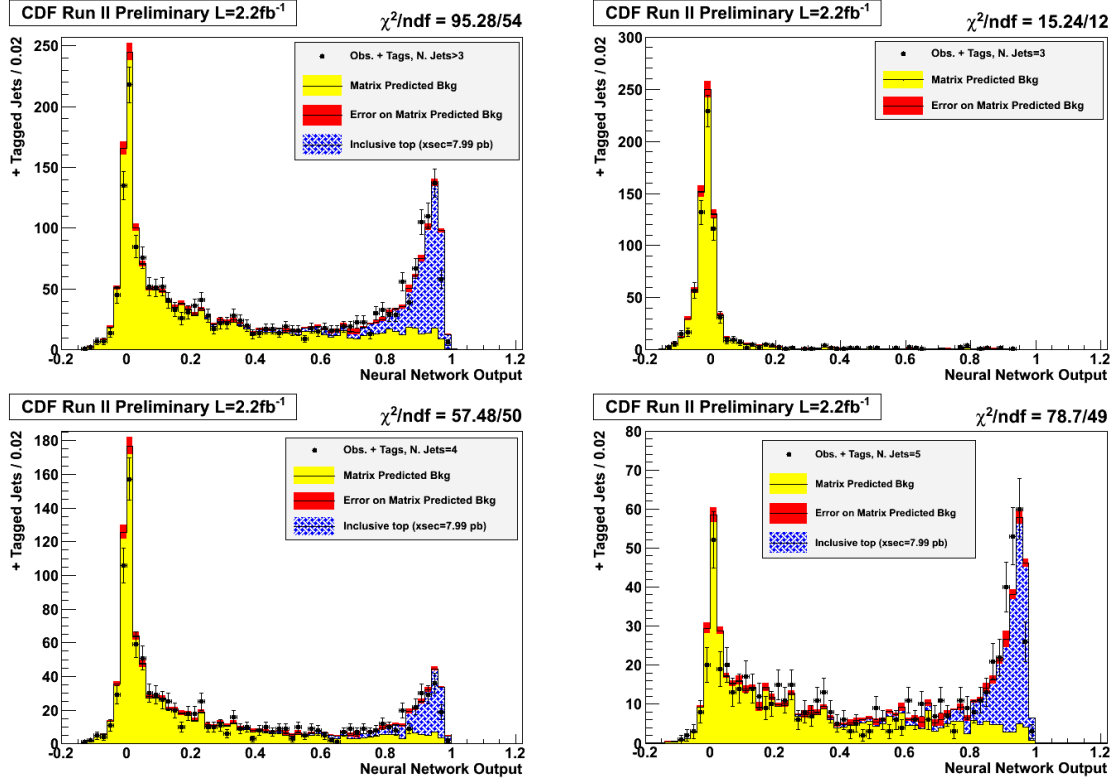
FIG. 2: Observed and matrix predicted number of positive tagged jets by Neural Network output in the multijet data for all events with at least 4 jets after prerequisites and for events with 3, 4 and 5 jets after prerequisites.

CDF Run II Preliminary, L=2.2 fb^{-1}

Source	Uncertainty
ϵ_{kin} systematics	
Generator dependence	3.9 %
PDFs	1.2 %
ISR/FSR	2.7 %
Color Reconnection	4.3 %
Jet Energy Scale	4.2 %
Trigger simulation	3.0 %
Primary Vertex $Z0$	0.2 %
ϵ_{tag} systematics	
SecVtX scale factor	3.9 %
Tagging matrix systematics	
Data control sample	2.5 %
Luminosity systematics	
Luminosity measurement	5.8 %

TABLE III: Summary of systematic uncertainties.

VI.   RESULTS

A.   Neural network cut optimization

Once we have trained the NN, we need to choose an optimized cut on its output to select the final data sample. The optimization procedure we seek is aimed at minimizing the statistical uncertainty on the cross section measurement,

CDF Run II Preliminary, L=2.2 fb^{-1}

N. of jets	3	4	5	6	7	8	9	Tot.
all hadronic (%)	0.12	0.46	1.66	4.86	7.79	10.18	9.88	2.29
e +jets (%)	26.64	25.19	35.46	36.20	35.10	33.60	33.83	32.08
μ +jets (%)	32.16	32.51	19.09	15.78	14.46	16.08	12.87	22.71
dileptonic (%)	6.46	2.30	1.07	0.70	0.52	0.64	0.00	1.45
had. τ +jets (%)	15.79	21.86	30.53	31.49	31.85	29.86	34.43	27.73
lep. τ +jets (%)	11.75	13.18	9.89	9.14	8.92	8.12	8.38	10.76
$\tau\tau$ (%)	1.06	1.15	0.65	0.47	0.43	0.44	0.30	0.77
$e/\mu + \tau$ (%)	6.03	3.36	1.64	1.35	0.92	1.08	0.30	2.16

TABLE IV: Expected sample composition after cut on Neural Network output greater than 0.8.

optimizing the expected number of b-tags over the background prediction. The former quantity is evaluated from inclusive Monte Carlo $t\bar{t}$ sample, the latter is derived from the b-tagging matrix application to data.

We decide to cut at $NNout \geq 0.8$, which gives the lowest expected statistical error on the cross section and an expected S/B ratio in terms of positive tags of 4.

Different choices of optimization of the cut have also been explored: methods to select a cut aimed at minimizing the total expected error on the cross section (statistical plus systematic) or maximizing the signal significance give results close to a cut on $NNout \geq 0.8$.

The summary of all the sources of systematic uncertainties to the cross section evaluation is listed in Tab. III for the chosen cut $NNout \geq 0.8$.

The expected sample composition after this cut is shown in Tab. IV.

B. Cross section measurement

After our selection we are left with 1420 events with at least 4 tight jets and we observe 636 positive b-tags. Observed and expected positive b-tags after selection for different jet multiplicities are shown in Fig. 3 along with the expected contribution of inclusive $t\bar{t}$ signal.
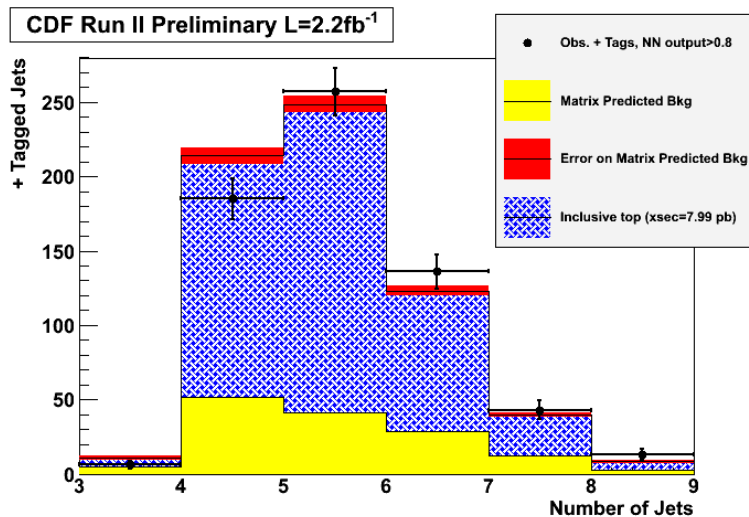

FIG. 3: Observed and matrix predicted number of positive tagged jets by jet multiplicity in the multijet data after cut on Neural Network output greater than 0.8, along with the expected contribution of inclusive $t\bar{t}$ signal.

Inserting in Eq. 2 the input parameters quoted in Tab.V, the measured cross section value is

$$\sigma_{t\bar{t}} = 7.99 \pm 0.55 \text{ (stat)} \pm 0.76 \text{ (syst)} \pm 0.46 \text{ (lumi)} \, pb = 7.99 \pm 1.05 \, pb$$

CDF Run II Preliminary, L=2.2 fb^{-1}

Variable	Symbol	Value
Integrated Luminosity (pb^{-1})	$\mathcal{L}$	2207.5 ± 128
Observed Tags	N_{obs}	636
Expected Background Tags	N_{exp}^{corr}	131 ± 9.6
Kin. efficiency (%)	ϵ_{kin}	3.53 ± 0.29
Ave. b-tagging efficiency	ϵ_{tag}^{ave}	0.811 ± 0.032

TABLE V: Input values for the cross section measurement.

Acknowledgments

[1] F. Abe, et al., Nucl. Instrum. Methods Phys. Res. A **271**, 387 (1988); D. Amidei, et al., Nucl. Instum. Methods Phys. Res. A **350**, 73 (1994); F. Abe, et al., Phys. Rev. D **52**, 4784 (1995); P. Azzi, et al., Nucl. Instrum. Methods Phys. Res. A **360**, 137 (1995); The CDFII Detector Technical Design Report, Fermilab-Pub-96/390-E

[2] T. Sjostrand et al., High-Energy-Physics Event Generation with PYTHIA 6.1, Comput. Phys. Commun. **135**, 238 (2001).

[3] G. Corcella et al., HERWIG 6: An Event Generator for Hadron Emission Reactions with Interfering Gluons (including supersymmetric processes), JHEP **01**, 10 (2001).

[4] The CDF coordinate system uses θ and ϕ as the polar and azimuthal angles respectively, defined with respect to the proton beam axis direction, z. The pseudo-rapidity η is defined as $\eta \equiv -\ln[\tan(\theta/2)]$. The transverse momentum of a particle is $p_T = p \sin\theta$ and the transverse energy is defined as $E_T = E \sin\theta$.

[5] T. Aaltonen et al., Phys. Rev. D 79, 052007 (2009)

[6] T. Aaltonen et al., Phys. Rev. D 76, 072009 (2007)

[7] M. Cacciari, et al., The $t\bar{t}$ Cross-section at 1.8 TeV and 1.96 TeV: A Study of the Systematics due to Parton Densities and Scale Dependence, JHEP **404**, 68 (2004).

[8] D. Acosta, et al., (CDF Collaboration), Phys. Rev D **71**, 052003 (2005).

[9] http://root.cern.ch/drupal/